You Get What You Measure

INTERNET PERFORMANCE MEASUREMENT AS A POLICY TOOL

Richard Bennett

NOVEMBER 2017

AMERICAN ENTERPRISE INSTITUTE

Executive Summary

The Federal Communications Commission (FCC) justified its 2015 reclassification of internet service providers (ISPs) as common carriers with a novel "virtuous circle" hypothesis. The agency correctly observed that broadband networks enable innovation in platform services such as Facebook, Amazon, and Netflix. It then speculated that platforms cause improvements in network performance through a feedback process.

The speculation provided the agency with a legal rationale for constraining internet services to its predetermined model: Unless internet services fit the agency's paradigm, platform innovation would not occur, and there would be no pressure on broadband networks to improve.

Meticulous measurements of broadband performance on the part of regulators and private firms confirm that speeds have improved at a 35 percent annual rate for the past decade. Contrary to the virtuous circle hypothesis, web speed—the time it takes for webpages to load—has only modestly progressed overall and has even regressed since the FCC's 2015 action.

Web performance has not been subject to the level of scrutiny focused in broadband platforms. This is due in part to the difficulty in measuring web performance. But the FCC has covertly politicized performance measurements. Its Measuring Broadband America (MBA) reports examine both webpage load times and broadband speeds but fail to analyze web data properly.

The load time of webpages does not improve over broadband networks faster than 12–15 megabits per second (Mbps). Early MBA reports reported this fact correctly, but those issued after the FCC redefined "broadband" to 25 Mbps have claimed a threshold value of 25 Mbps even though the underlying data have not changed.

The emphasis on one facet of internet performance, such as last-mile broadband networks, tends to minimize other factors that may be more important to the user, such as the performance impact of tracking networks, browsers, webpage design, and web server performance. In addition, relying on active measurement tools creates opportunities for gaming the system that are not possible in passive systems that merely observe application and network events in real time.

This paper explores opportunities for developing performance tools more responsive to the broader social goal of better end-to-end internet performance across the broad span of applications. It finds that systems for capturing passive measurements and sharing them among ISPs, web developers, and other responsible parties may be useful for accelerating the web experience.

The performance of websites over time is a neglected facet of internet measurement that deserves more attention. In an era of increasing internet consolidation, smaller sites and platforms are squeezed by larger competitors able to invest in private infrastructure. Better insight into web performance and increased flexibility in contracts between content platforms and broadband platforms may mitigate investment inequality effects.

You Get What You Measure

INTERNET PERFORMANCE MEASUREMENT AS A POLICY TOOL

Richard Bennett

The research literature on internet performance measurement is quite rich.[1] Surveys of measurement tools such as “A Study of Traffic Management Detection Methods & Tools”[2] and “A Survey on Internet Performance Measurement Platforms and Related Standardization Efforts”[3] describe a multitude of tools such as NetPolice, NANO, DiffProbe, Glasnost, ShaperProbe, Chkdiff, SamKnows, BISmark, Dasu, Netradar, Portolan, RIPE Atlas, and perfSONAR intended for use in detecting net neutrality violations.

In addition to tools developed for academic research and policy enforcement, internet users rely on Speedtest and OpenSignal for troubleshooting. Finally, proprietary systems such as those developed by Akamai,[4] Sandvine,[5] Ookla,[6] and Cisco[7] are used to compile “State of the Internet” analyses aggregating several views of the internet.

While current tools are quite useful for measuring the performance of broadband networks, they are much less useful for examining how well the internet operates as a whole. The internet is an “end-to-end network of networks” in which performance depends on a series of cooperating networks and network-attached devices and services.[8]

From the user perspective, the web appears to be slowing down.[9] While this trend has become received wisdom, traditional measures of broadband performance continue to show improvement: Akamai’s measurements of “average peak connection speed” show US average speed increased an average of 29 percent per year between 2010 and 2017, and the Federal Communication Commission (FCC) reports a 43 percent average annual increase from 2011 to 2015.[10]

The emphasis on one facet of internet performance, such as last-mile broadband networks, tends to minimize other factors that may be more important to the user, such as browsers, webpage design, and web server performance. In addition, relying on active measurement tools creates opportunities for gaming the system that are not possible in passive systems that merely observe application and network events in real time. However, passive systems have privacy issues.

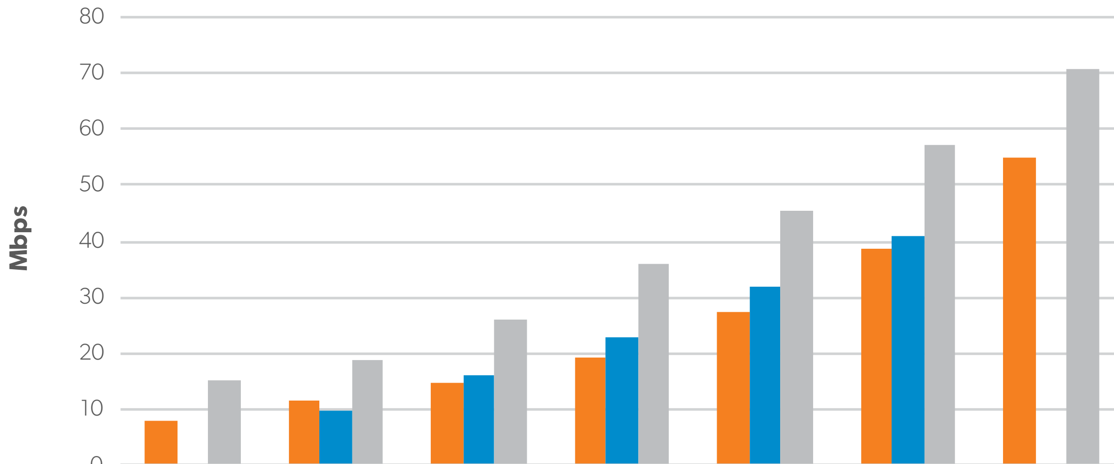
This paper explores opportunities for developing performance tools more responsive to the broader social goal of better end-to-end internet performance across the broad span of applications. It finds that a system for capturing passive measurements and sharing them between among internet service providers (ISPs), web developers, and other responsible parties may be useful for accelerating the web experience.

Measuring Broadband Networks

Wired networks in the US and the rest of the world have become dramatically faster since becoming broadly subject to public measurement circa 2010 (Figure 1).

From third quarter 2010 to third quarter 2016, Akamai’s “State of the Internet” reports show an increase in “average peak connection speed” from

Figure 1. Average US Wireline Download Speeds, 2010–16



	2010	2011	2012	2013	2014	2015	2016
Ookla	8	11	15	19	27	39	55
FCC		10	16	23	32	41	
Akamai	15	19	26	36	45	57	71

Sources: Akamai, "State of the Internet," 2010–16, https://www.akamai.com/us/en/about/our-thinking/state-of-the-internet-report/global-state-of-the-internet-connectivity-reports.jsp; Federal Communications Commission, Office of Engineering and Technology and Consumer and Governmental Affairs Bureau, "Measuring Broadband America," 2011–16, https://www.fcc.gov/general/measuring-broadband-america; and Speedtest by Ookla, "United States Speedtest Market Report," 2010-16, http://www.speedtest.net/reports/united-states/.

15.4 megabits per second (Mbps) to 70.9 Mbps across wireline broadband networks in the US, an annual average of 29 percent, and an overall increase of 460 percent.[11] The global average is even more impressive, jumping nearly 700 percent since 2010.

SamKnows tests of the broadband "promise index" (actual download speed as a percentage of advertised speed) show steady improvement. The FCC's SamKnows-based "Measuring Broadband America" reports, conducted annually since 2011, show the overall US promise index improved from 87 percent in 2011 to 101 percent in 2014.[12]

Once delivered speeds exceeded advertised ones, the FCC stopped reporting overall promise index averages. In their place, the commission shifted to median broadband speeds averaged across all providers and plans.[13]

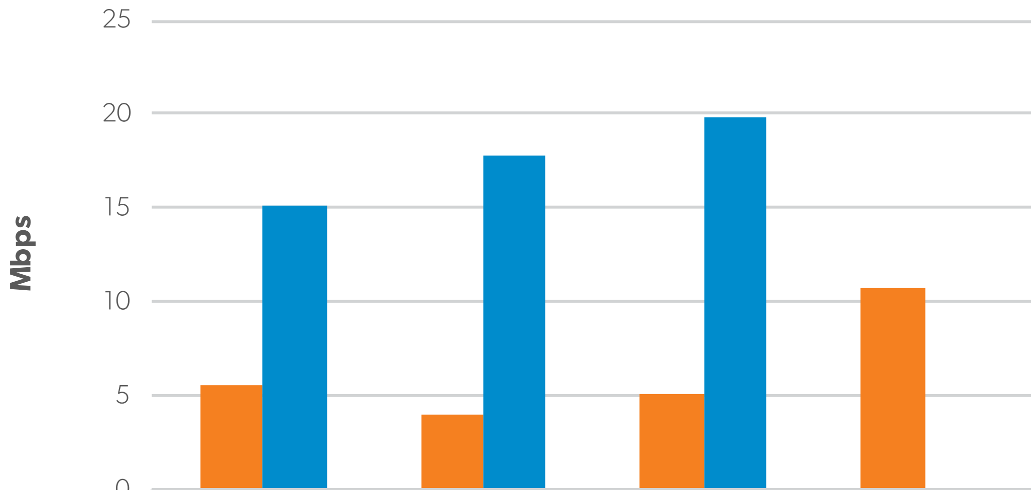
By the FCC's estimation, US median download speed increased from 10 Mbps in 2011 to 41 Mbps in 2015. The average annual speed increase was 43 percent over this period, with a total increase of 410 percent.[14]

Ookla measurements show US broadband download speed increasing from 7.8 Mbps in August 2010 to 70.6 Mbps in July 2017 for an average annual gain of 40 percent and an overall increase of 900 percent.[15]

Mobile speeds are more variable and harder to measure than fixed broadband. Regulators in the Americas, Europe, and East Asia use SamKnows for accurate, representative measurements of wired networks. But SamKnows depends on a special internet gateway known as the Whitebox for continuous measurement; such a device is neither practical nor useful for measuring mobile performance.

Akamai has not always meaningfully aggregated performance data for national measurement. Before 2014, Akamai reported mobile performance data on an anonymized carrier-by-carrier basis. This form of reporting provided a range of values with no

Figure 2. Average US Mobile Download Speeds, 2014–17



	Q1 2014	Q1 2015	Q1 2016	Q1 2017
Average	5.5	4	5.1	10.7
Peak	15.1	17.8	19.8	

Source: Akamai, "State of the Internet," 2014–17, https://www.akamai.com/us/en/about/our-thinking/state-of-the-internet-report/global-state-of-the-internet-connectivity-reports.jsp.

possibility of weighting to produce a national aggregate. From quarter to quarter, these measurements did not even include the same number of carriers. The Akamai data also dropped peak averages in 2017, so we are left without a meaningful range of reliable data for mobile broadband (Figure 2).

This limited data suggest a year-on-year average increase of 44 percent for average connection speed—taken with multiple Transmission Control Protocol (TCP) virtual circuits active—and a 15 percent average annual increase of peak connection speeds.

These speeds are substantially slower than wired broadband speeds: Compare the 19.8 Mbps for mobile in first quarter 2016 with the corresponding figure for wired, 67.8 Mbps. The rate of increase for wired broadband over the period from 2014 to 2016 was also higher than for mobile: 25 percent versus 15 percent.

Ookla restored international speed comparisons in August 2017, but its data include only a year of history. For completeness, see Figure 3 for Ookla's US data charts for fixed and mobile. Note that Ookla's figures coincide closely with Akamai's average peak connection speeds, especially for mobile. In 2016, Akamai measured an average peak of 19.8 Mbps for mobile, while Ookla measured 19.3 Mbps in July and 22.2 Mbps in December.

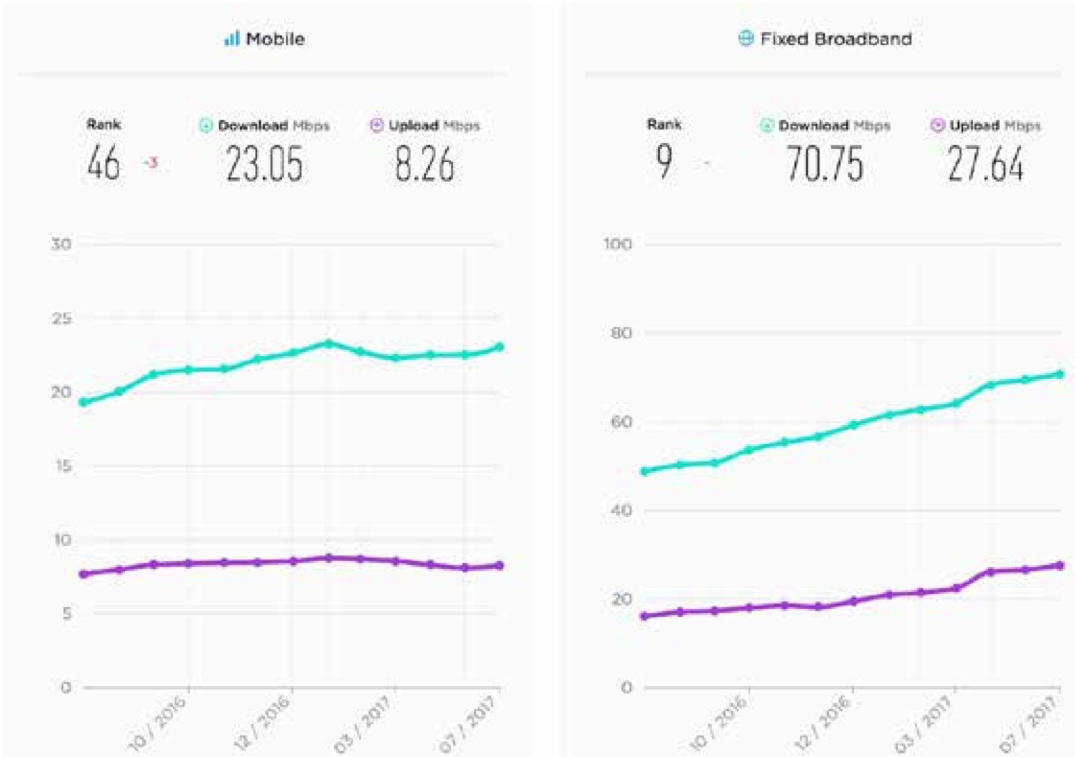
Akamai's ranking of wired broadband average connection speed in the US (10th) coincides closely with Ookla's ranking (9th).

The Ookla measurements show that the average speed of mobile broadband in the US is close to the FCC's 25 Mbps broadband benchmark on the download side and well above the 3 Mbps benchmark on the upload side.

Measuring the Web

But end user perceptions of the web do not mirror empirical improvement in broadband performance. While there is not a database of webpage load times nearly as comprehensive as the broadband

Figure 3. US Broadband Speeds, 2016–17



Source: Speedtest by Ookla, "United States Speedtest Market Report," August 3, 2016, http://www.speedtest.net/reports/united-states/.

measurements Akamai, Ookla, and SamKnows conducted, the perception of a slower web is widespread. Consider the following anecdotes.

In Jim Rapoza's "The Web Is Getting Slower" he states:

> One thing you can always rely on technology to do is speed things up. Everything, from processors to phones to networks gets faster. Heck, there are actual laws that define this phenomenon. So when at a recent Akamai analyst event a speaker made the off-hand comment that the Web is getting slower, it pretty much made me sit up in my seat and say "what?"
>
> My first gut instinct was to say "No way, this is technology, things don't get slower. I used to have a modem, now I have fibre. I used to use a WAP browser for mobile web, now I have fast 4G and LTE connections." But once that initial instinct passed, I had to admit, it sure did seem that many of my recent web browsing experiences were less than satisfactory from a performance standpoint.
>
> So what's causing this slowdown? Is it the result of problems in the core of the Internet's infrastructure? Well, while there have been cases of hardware problems causing Web slowdowns, as well as performance issues caused by political fights between major carriers and streaming video providers, the cause of the Web's slowdown is actually coming from the other side of the infrastructure.[16]

In Kalev Leetaru's "Why the Web Is So Slow and What It Tells Us About the Future of Online Journalism," he states:

> While there are certainly those philosophically opposed to online advertising and those who dislike

> the intrusiveness of some ads, one of the biggest driving forces behind ad blockers has been the immense degradation modern advertising practices can cause to user experience, slowing webpages down by burdening web browsers with hundreds or even thousands of requests to fully load all of the advertisements and analytics.
>
> A brief look at a cross-section of major news websites from throughout the world suggests that news websites are among the most bloated sites on the web today, with one tested site making upwards of 6,500 distinct requests to over 130 different domains just to display its homepage. To stem the tide of ad blocking, the journalism industry should rethink the design of news websites and actively explore open source mobile optimization platforms like Google's Accelerated Mobile Pages (AMP).[17]

In Hope King's "The Web Is Getting Slower," she states:

> The spinning wheel of death never seems to stop turning these days.
>
> It's not you. Web pages really are loading slower.
>
> The average site is now 2.1 MB in size—two times larger than the average site from three years ago, according to data tracked by HTTP Archive.
>
> There are a few reasons for this added weight.
>
> Websites are adding more attention-attracting videos, images, interactivity plug-ins (comments and feeds) and other code and script-heavy features that clog up broadband pipes and wireless spectrum.
>
> Sites also have ramped up their usage of tracking and analysis tools to learn more about their visitors. Inserting third-party data trackers not only increases a website's weight, but also the number of separate data fetching tasks, which leads to slower load times as well.[18]

The limited data available on web performance are consistent with the perception that the web is slowing down, or at least not improving at the same rate as networks.

Akamai Web Measurements

Akamai measured webpage load times from the third quarter of 2013 until the fourth quarter of 2016, when it withdrew the test to rework it. Akamai's web measurement tool—known as real user measurement (RUM)—embeds JavaScript in webpages to assess and report anonymized load time data to a monitor.[19]

Unlike the synthetic testing approach SamKnows and Ookla use for broadband speeds, RUM is passive, an attempt to capture real-life data in the same way that Akamai's connection speed measurements of TCP performance do.[20]

Akamai's connection speed measurements are taken on servers in the company's content delivery network, while RUM measurements are taken by browsers executing Akamai's JavaScript code.

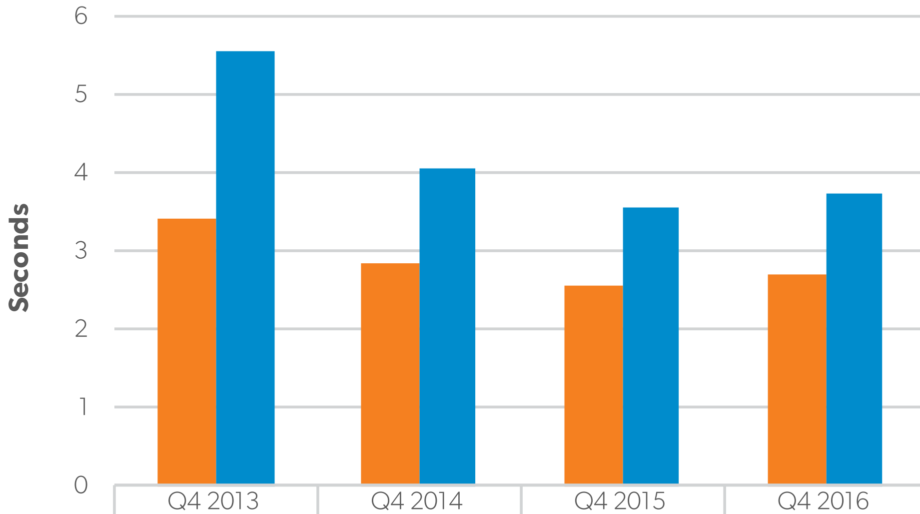
Measured results were not promising: Average page load time for wired browsing in the US was 3.4 seconds in fourth quarter 2013 and 2.71 seconds in fourth quarter 2016. While this is an improvement of 25 percent, wired broadband speed increased by 88 percent over this period, from 43 to 81 Mbps[21] (Figure 4). Because Akamai's measurements are taken on pages hosted on the company's content delivery network (CDN), they do not reflect the performance of unaccelerated pages.

WebPageTest.org Measurements

HTTP Archive and WebPageTest provide an archive of website loading tests extending back to November 2010. Cataloged tests are one-offs that may not represent the website's behavior over a period of time. These tests are presented in video format, so we can watch text and images appear.

If this archive demonstrates anything at all, the takeaway would be that website operators are able to control their performance. Snapshots of four major newspaper websites show three becoming faster since 2010 and one, the *New York Times*, growing slower. Of the three that are now faster, the *Washington Post* shows the greatest improvement, cutting the time

Figure 4. US Webpage Load Time, 2013–16



	Q4 2013	Q4 2014	Q4 2015	Q4 2016
Fixed	3.4	2.9	2.6	2.7
Mobile	5.6	4.1	3.6	3.8

Source: Akamai, "State of the Internet," 2013–16, https://www.akamai.com/us/en/about/our-thinking/state-of-the-internet-report/global-state-of-the-internet-connectivity-reports.jsp.

to begin showing text above the fold in two seconds, down from six in 2010.

Two of the news sites are reasonably speedy: The *Washington Post* and the *Los Angeles Times* begin to populate their main pages with text in two seconds (Figures 5–6).

But the New York papers are tragically slow, showing nothing visual for five seconds (Figures 7–8). This pattern of slow load time is nothing new for either site, but performance appears to be degrading at the *New York Times* as above-the-fold text that used to appear in four seconds in 2010 now takes six seconds.

Webpage load time has major implications for e-commerce, so sales-oriented sites tend to be quicker to load than news sites. Amazon and eBay begin to show text above the fold in 1.5 seconds, and both sites are faster than they were in 2010. Not surprisingly, Amazon outperforms eBay in web load time and in growth, market share, and profit (Figures 9–10).

Traffic Management Detection Tools

Surveys of measurement tools such as "A Study of Traffic Management Detection Methods & Tools"[22] and "A Survey on Internet Performance Measurement Platforms and Related Standardization Efforts"[23] describe a multitude of tools such as NetPolice, NANO, DiffProbe, Glasnost, ShaperProbe, Chkdiff, SamKnows, BISmark, Dasu, Netradar, Portolan, RIPE Atlas, and perfSONAR designed to detect net neutrality violations.

These tools have neither a proper vantage point nor ground truth knowledge of the behavior they seek to detect. Lucy Hazell, Peter Thompson, and Neil Davies conclude such approaches have limited value:

> None of the TMD methods studied satisfy all the key attributes that would make them suitable for effective practical use. In particular, those that are currently in active deployment generate significant volumes of traffic, which would

Figure 5. ***Washington Post*** **Website Animation**

Source: WebPageTest, "Test a Website's Performance," WebPageTest.org.

Figure 6. ***Los Angeles Times*** **Website Animation**

Source: WebPageTest, "Test a Website's Performance," WebPageTest.org.

Figure 7. ***New York Times*** **Website Animation**

Source: WebPageTest, "Test a Website's Performance," WebPageTest.org.

risk damaging the QoE [quality of experience] of other users if applied widely, and incur costs to the service providers of carrying this traffic; thus they may be unsuitable for large-scale use. The reliability of these tools would require further study, using a uniform test environment in which their performance could be objectively compared. . . .

Finally, these tools are limited in that they aim only to detect the presence of differential (intra-user) traffic management, as the detection of non-differential traffic management (inter-user or aggregate) was not their goal.

These tools are not sufficient to enable effective detection and location of TM application along a

Figure 8. ***Wall Street Journal*** **Website Animation**

Source: WebPageTest, "Test a Website's Performance," WebPageTest.org.

Figure 9. Amazon Website Animation

Source: WebPageTest, "Test a Website's Performance," WebPageTest.org.

Figure 10. eBay Website Animation

Source: WebPageTest, "Test a Website's Performance," WebPageTest.org.

fragmented digital delivery chain such as that in the UK. Our conclusion is thus that no tool or combination of tools currently available is suitable for effective practical use.[24]

More work needs to be done in the area of differential treatment identification; such work depends heavily on the vantage point of the detection instrument and the nature of the differential treatment it seeks to detect.

Rationales for High Broadband Speeds

Conventional wisdom holds that broadband speed is a significant factor in webpage load time only at low levels of performance. The first four "Measuring Broadband America" reports declared the sweet spot for webpage load time to be 10 Mbps. This is the statement from the 2011 report:

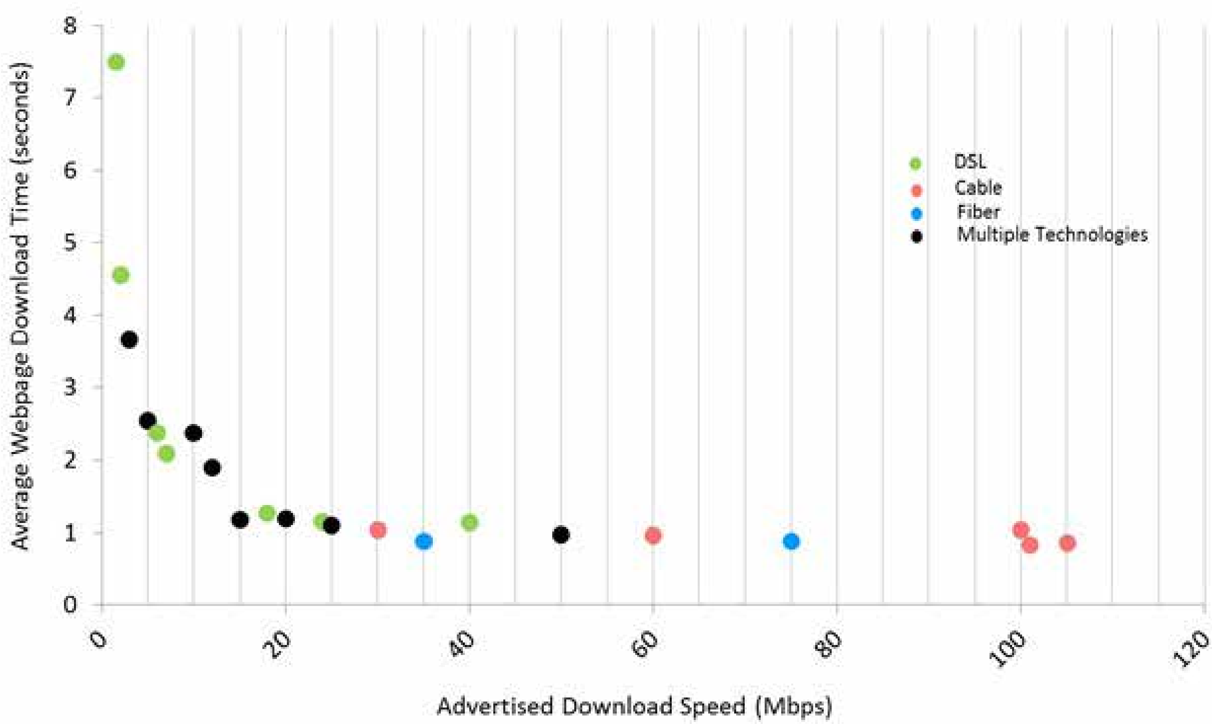
> In specific tests designed to mimic basic web browsing—accessing a series of web pages, but not streaming video or using video chat sites or applications—performance increased with higher speeds, but only up to about 10 Mbps. Latency and other factors limited performance at the highest speed tiers. For these high speed tiers, consumers are unlikely to experience much if any improvement in basic web browsing from increased speed—i.e., moving from a 10 Mbps broadband offering to a 25 Mbps offering.[25]

Subsequent reports contained similar declarations about a 10 Mbps plateau point until the 2015 report revised the threshold to 25 Mbps:

> Users subscribing to a service tier with a 1.5 Mbps download speed on average wait for approximately 7.5 seconds for a webpage containing text and images; users subscribing to a service tier with a 5 Mbps download speed on average wait only approximately 2.5 seconds; and users subscribing to a service tier with a 25 Mbps download speed on average wait on only approximately 1 second. Subscribers to service tiers with an advertised download speed exceeding 25 Mbps on average do not experience significantly reduced webpage download time.[26]

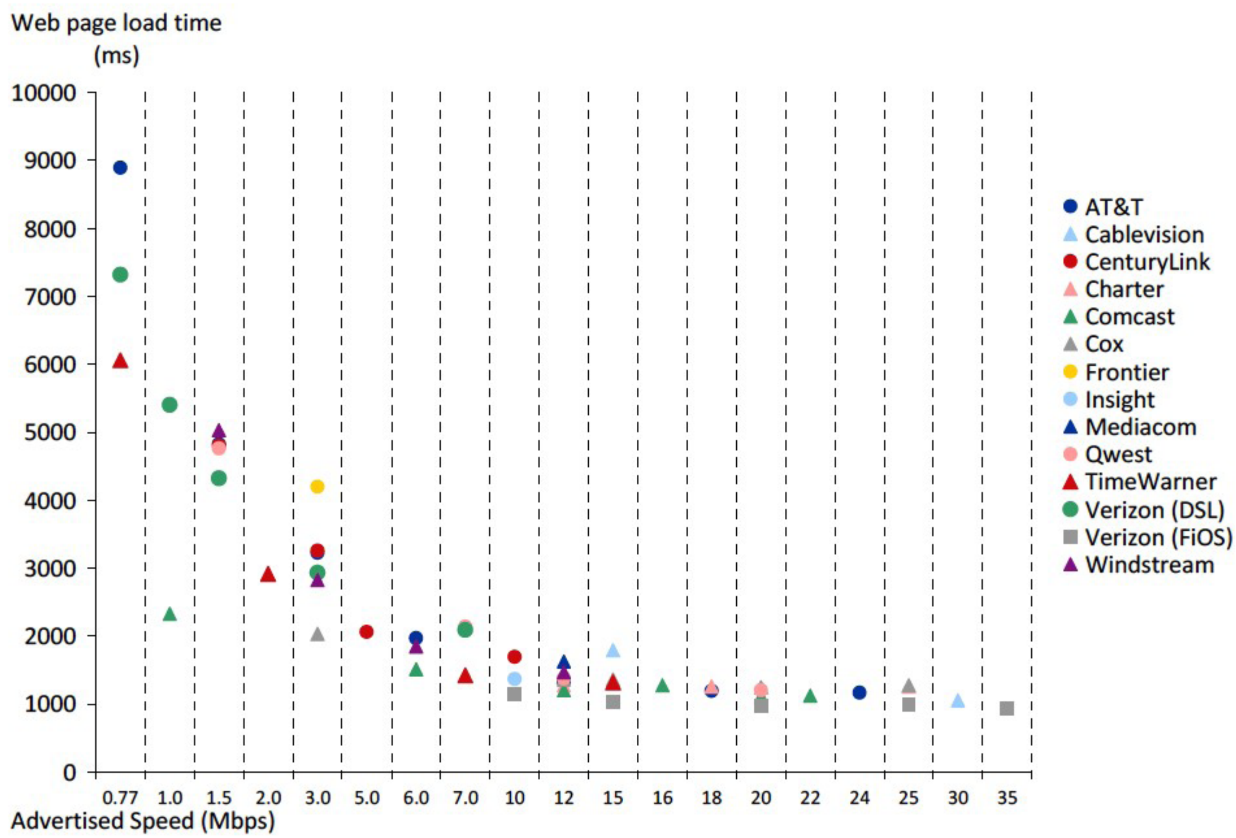
The 2015 report includes a chart (Figure 11) suggesting the plateau is actually 15 Mbps. The report's 25 Mbps analysis is narrowly correct but misleading.

Figure 11. Webpage Load Time as a Function of Broadband Speed, 2014



Source: Federal Communications Commission, Office of Engineering and Technology and Consumer and Governmental Affairs Bureau, "Measuring Broadband America Fixed Broadband Report—2015," December 30, 2015, https://www.fcc.gov/reports-research/reports/measuring-broadband-america/measuring-broadband-america-2015.

Figure 12. Webpage Load Time as a Function of Broadband Speed, 2015



Source: Federal Communications Commission, Office of Engineering and Technology and Consumer and Governmental Affairs Bureau, "Measuring Broadband America," 2016, https://www.fcc.gov/general/measuring-broadband-america.

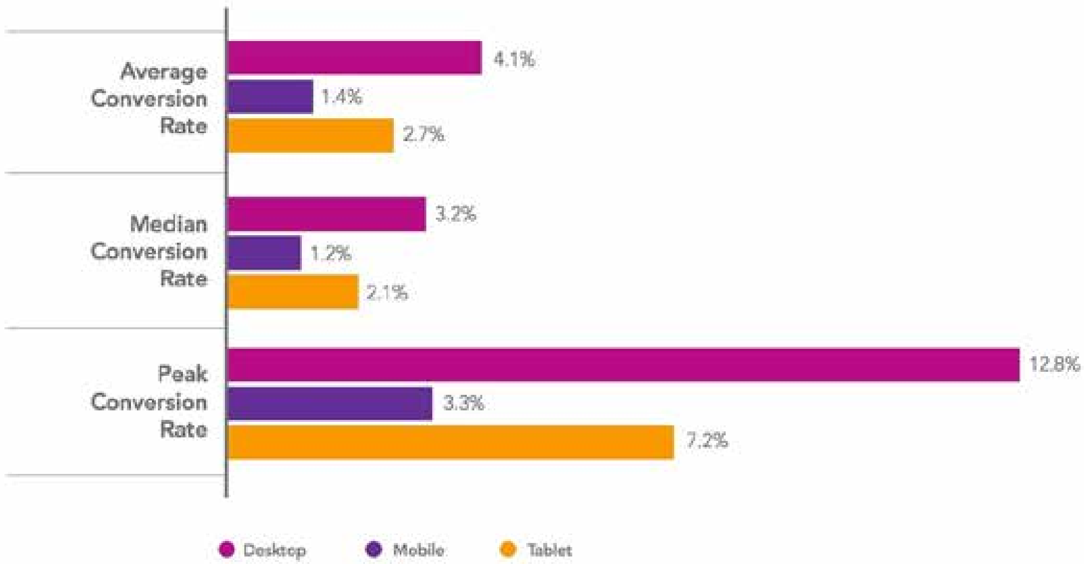
In fact, little changed between the data on which the 2015 report was based and the data behind the original 2011 report. The plateau for web speed is reached on broadband networks running at 12–15 Mbps. Figure 12 demonstrates this fact.

In January 2015 the FCC had raised the threshold definition of broadband from 4 Mbps to 25 Mbps:

> Based on the record, we find that a 25 Mbps/3 Mbps benchmark reflects "advanced" telecommunications capability. We have recognized that the concept of broadband does not stand still, but instead must evolve and after a new and updated review of the market, we find that a speed benchmark of 25 Mbps/3 Mbps best captures the statutory definition envisioned by Congress.[27]

The FCC's reasoning on the redefinition is thin, citing a policy statement by the City of Boston and miscalculations about the needs of popular applications, but it is most persuasive in arguing that "setting a benchmark at 25 Mbps/3 Mbps may allow us to retain the same speed benchmark for multiple years." The text of the 2015 "Measuring Broadband America" report tracked the new definition even though the data did not.

Figure 13. Conversion Rates by Device Type and Load Time



Source: Akamai, "The State of Online Retail Performance," 2017, https://www.soasta.com/wp-content/uploads/2017/04/State-of-Online-Retail-Performance-Spring-2017.pdf.

Rationales for Fast Webpage Load Times

It would be desirable to have websites that loaded pages in one second or less. Human factors research suggests that sub-second load times are consistent with maintaining concentration while clicking links and navigating from page to page. Robert Miller's classic paper on computer human interfaces maintained that humans expect delays of one to two seconds for nontrivial tasks and one- to two-tenths of a second for trivial ones.[28]

Akamai released a report in the spring that compiled passive measurements of webpages—using the RUM methodology—to provide insight about performance factors as they relate to e-commerce.[29] This research suggests that the rate of conversion from shoppers to buyers is highest when page load time is between 1.8 and 2.7 across devices: 1.8 seconds on desktop devices and 2.7 seconds on mobile devices other than tablets[30] (Figure 13).

E-commerce has a clear incentive to maintain short load times: It seeks to make sales, and potential customers are discouraged by waiting for pages to load. Researchers also track *bounce rates* of e-commerce sites. In contrast to conversions (sessions in which the user makes a purchase, signs up, or otherwise fulfills the site's goal), bounces are sessions in which the user immediately leaves the site without going past the first page.

Slow load times affect bounce rate more than conversion rate: Optimal bounce rates require load times of 700 milliseconds for desktops and 1.2 seconds for mobile.[31] For e-commerce, website performance has a direct relationship with revenue.

General websites also experience revenue effects from slow page loads, but they are somewhat less direct. The revenue model for the web generally depends on ad sales, but ads are principal drivers of poor performance. The more ads a page carries, the higher its revenue potential when all other factors—such as traffic and click-throughs—are equal, but slow page loads encourage bounces.

According to HTTP Archive, the typical webpage now accesses 19 domains and forms 34 TCP connections.[32] Some websites are substantially higher than average; the Ghostery plugin reports 51 connections to trackers, analytics, and social media for the *New York Times* front page, 41 for the *Washington Post* and *Wall Street Journal*, and 81 for the *Los Angeles Times*.[33] Many of these trackers are flagged as "slow or insecure," meaning they impair load time or fail to protect privacy.[34]

So commercial websites are caught in a vise between raising revenue by carrying ads and increasing traffic by loading fast. Site operators therefore have an interest in knowing which ads and which trackers make the greatest contribution to revenue while imposing the least burden on page load time.

Technical Factors Affecting Webpage Load Time

The most obvious factors affecting webpage load time are raw page size and broadband bandwidth. The average webpage is just over 3 MB (24 megabits) in total size, and the average wired broadband pipe offers 87 Mbps.[35] Hence, the average page load time should be no more than 275 milliseconds. In reality, webpages load in 1.3 seconds (FCC), 2.6 seconds (Akamai), or five seconds (Pingdom).[36]

Many factors can affect webpage load time:

1. **Web Server Performance.** Individual web servers can service a relatively small number of concurrent accesses without noticeable slowdown. The popular Apache web server has a default limit of 256 for the "MaxRequestWorkers" parameter governing concurrent accesses. Each access requires using finite memory, central processing unit (CPU) cycles, and disk resources. High-capacity websites run on CDNs and use sophisticated load balancing to make the most of total capacity, but redirecting users to far-away servers affects performance.

2. **Browser Performance.** Browsers are designed to optimize load time, but not all optimizations are ideal for all pages. Common optimization tricks are multiple concurrent TCP sessions, displaying page elements before they are completely loaded, caching, and prediction. Some browsers have reduced their concurrent TCP streams over time upon discovering that too many TCP streams reduce performance. With so many page elements coming from domains external to the main page, even this simple parameter is hard to optimize because external sites do not load at the same rates.

3. **Webpage Design.** This includes the order of loading style sheets, main page text, images, and other objects, as well as design choices related to revenue. Generally, webpage designers strive to load the portions of pages immediately visible to users quickly and load other page elements on more relaxed schedules. Bounces cost sites money if users leave before ads are loaded. Webpage design has to balance aesthetics, performance, and revenue.

4. **PC and Smartphone Performance.** Every page stresses CPU, memory, and storage on the user device and server. Some pages are oddly CPU intensive, and some browsers achieve fast loading by using as much CPU and memory as possible. Like concurrent streams, there is a practical limit to CPU consumption. Ad blockers and virtual private networks increase CPU load; hence, they hurt load time.

5. **TCP Performance.** TCP is a poor fit for the web because it has to manage internet congestion and assure end-to-end data integrity. Thanks to the Jacobson algorithm's "slow start" feature, new TCP connections are capable of less throughput than old ones.[37] Consequently, webpages that open many TCP connections, especially for sessions of short duration, cannot transmit at "wire speed." Attempts to rectify this issue at the application layer—such as HTTP/2—have not been completely successful, but experiments continue on smoothing TCP performance. Google's bottleneck bandwidth and round-trip propagation time (BBR) congestion control is the latest and currently most

hopeful.[38] BBR uses latency measurements to avoid congestion rather than packet loss.

6. **Network Congestion.** Congestion is endemic to the internet because the system lacks an effective method of flow control that applies back pressure to applications to match offered load to usable capacity at sub-second units of time. Traditionally, this problem has been addressed by TCP acting on its own, but this approach creates several inefficiencies such as a sawtooth traffic pattern, retransmissions, and failure to use all available capacity. The problem is solvable in theory even though it has not been solved in practice.

7. **Perception.** Webpage load time is ultimately a matter of human factors. Human users have expectations of the time tasks should take to perform, primarily triggered by visible conditions. Much of the delay in loading webpages comes about from page elements that are of low interest to users—such as ads—and other background activities such as analytics. A great deal of the emphasis on webpage load time is aimed at making the portions of pages above the fold appear to load first.

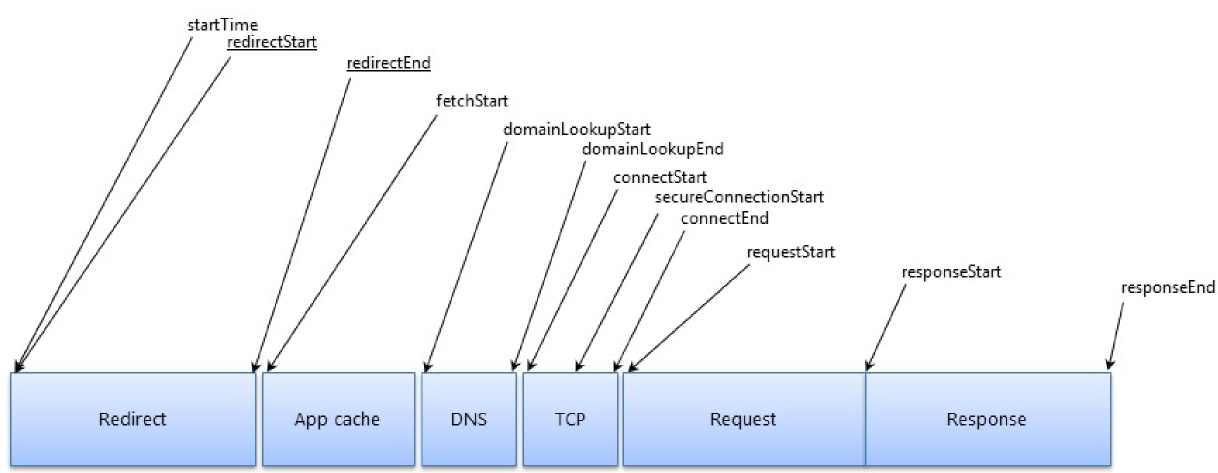
In the ideal scenario, users could detect the proximate causes for slow page loads and act to resolve them. Generally, users have access to only three forms of mediation: upgrading broadband connections, upgrading their desktop or smartphone environments, or complaining to the party responsible for the slowdown, typically the broadband or website operator.

The overemphasis on broadband performance probably helps spread the perception that the ISP is responsible for web performance. Troubleshooting articles of recent vintage focus on broadband performance as the key issue, failing to go beyond contrasting Wi-Fi performance with Ethernet performance.[39] While this is sometimes the case, it is likelier that the website itself is to blame for sluggish browsing.

Instrumenting Webpages for Performance Measurement

The World Wide Web Consortium (W3C) has been working on passive web performance measurement tools for some time and has made considerable progress in building performance measurement into the fabric of the web.[40]

Figure 14. Timing Attributes Defined by the W3C PerformanceResourceTiming Interface



Source: Arvind Jain et al., "Resource Timing Level 1," W3C Standards, March 30, 2017, https://www.w3.org/TR/resource-timing-1/.

The first part of this effort was to define high-resolution timers that can be accessed by other functions to accurately measure performance. This work is embodied in the High Resolution Time Level 2 standard.[41] Another standard, Resource Timing Level 1, allows web code to access detailed measurement of page events using the high-resolution timer.[42] The events in question include start of request for a page, initial page data, and page complete ("responseEnd").

W3C work also covers server timing, navigation, and the isolation of page elements. It is worth noting that web performance can be measured in the server, browser, or network. While a comprehensive view would combine the three vantage points, we are not there yet because we in some sense lack complete pictures in all of them. For the kind of accountability that both users and regulators would like to have, it will be necessary to advance the work and combine the results.

Of the three, it appears that browser-based measurement is the least advanced and most important at the moment. This has not always been the case, but with broadband speeds well above threshold levels and good instrumentation in at least the CDN-based servers, browsers have become a measurement bottleneck.

The continued development of standards-based metrics—and means for sharing these data with responsible parties—is vital to the assessment of the web in particular and the internet in general.

End-to-End Performance Measurement

To measure the real-world performance of any system, it is necessary to either monitor the system or create a realistic simulation. At its heart, internet infrastructure is statistical in nature: Internet transmission uses packet switching, a form of statistical sharing of communication networks. Web-based services are shared by multiple users, so performance measurements have to accommodate a variety of load scenarios ranging from single user to denial-of-service attacks in which the service falls over.

Simulating such a system is nearly as difficult as building it, and verifying a simulation requires ground truth measurements from the real system. Hence, meaningful internet performance measurement requires the measurement of real system operation.

End point devices probably offer the most meaningful vantage point, but they are prone to ambiguity unless their measurement code takes system factors into account. We have to ask questions about the device's performance limits and the state of its internal resources at the time when measurements are taken. For example, a CPU overloaded with computation tasks, an exhausted real memory pool, or a hyperactive disk drive can impair webpage transfer and rendering. But real systems do suffer from resource constraints that affect complex webpages more than simple ones, so resource status is a factor that affects real-world performance.

Web performance can also be measured by web servers, but only in a limited way. Webpages are assembled from page elements located on many servers—typically 19 per page and sometimes more than 100.[43] Each server can only measure the rate at which it pushes elements to its internet connection, not their transmission time or rendering time.

Finally, end-to-end internet interactions can be measured by network monitors close to the user device. This vantage point provides clarity on the performance characteristics of both client and server but tends to conflate network factors with server factors. So the most promising direction may be network monitors with code that distinguishes server factors from network factors. Several efforts are underway to develop such monitors.

PAIN: A Passive Web Speed Indicator for ISPs

Passive Indicator (PAIN) is a system researchers at the Politecnico di Torino devised that was presented at the ACM SIGCOMM workshop on "QoE-Based Analysis and Management of Data Communication Networks" in August 2017.[44] It is a quality of experience (QoE) tool targeted at ISPs because "ISPs are evaluated based on the experience of end-users while interacting with third-party services."[45] Its approach is to leverage passive flow and domain name system (DNS) measurements unaffected by encryption.

PAIN gathers passive TCP flow data and DNS measurements from traffic logs in ISP devices. Together, these logs enable a view of webpage construction to be assembled and performance to be measured from time stamps. The developers describe its central intuition:

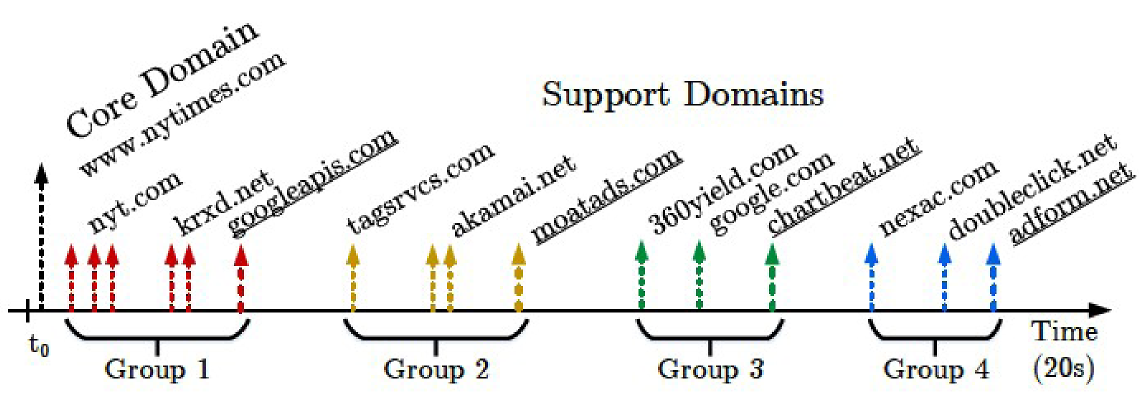
> Once users reach a website, their browsers open many flows to different servers to fetch HTML objects, scripts and media content. We call the Fully Qualified Domain Name (FQDN) associated with the first contacted server the Core Domain and the remaining contacted FQDNs the Support Domains. An example is provided in [Figure 15], which illustrates with colored arrows the moment in which flows to support domains appear after a visit to the core domain www:nytimes:com. Given core domains of interest, PAIN automatically learns contacted support domains, as well as the typical order in which such flows appear in the network, creating groups of support domains. In the example, PAIN learns 4 groups from the observed network traffic. PAIN then considers the delay to observe flows of each group a performance indicator. It uses visits to the website from all users to (i) observe probable patterns; (ii) identify checkpoints that model the download process; and (iii) measure the delay to pass checkpoints, i.e., automatically building a benchmark.

In essence, PAIN constructs a map of the references from the main body of a webpage's text to external domains. It then measures the page's performance by simply calculating elapsed time from the request for the page to the last external response from the page's support domains. It capitalizes on the facts that DNS requests are made in clear text and that remote internet protocol (IP) addresses returned by DNS are unprotected by encryption even when payloads are hidden.

Aggregating flow data for support domains are at least as useful as measuring flows from the core domain because support domains are probably less diverse than the set of core domains seen by the ISP across all users. The support domains for ad placement are not at all diverse as the services provided by Google, DoubleClick, and Facebook dominate the internet advertising market.

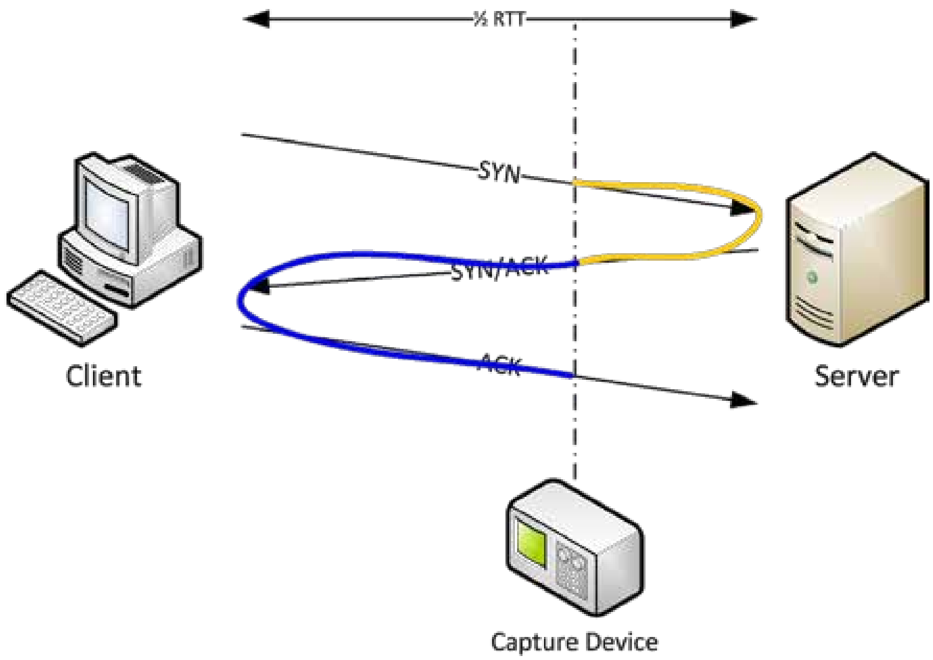
This approach is quite clever and reasonably accurate. The data are tremendously useful because they are able to characterize performance from support

Figure 15. PAIN Flows in a Nytimes.com Visit



Note: PAIN uses the time to contact support domains to monitor performance.
Source: Martino Trevisan, Idilio Drago, and Marco Mellia, "PAIN: A Passive Web Speed Indicator for ISPs," ACM SIGCOMM Workshop on QoE-based Analysis and Management of Data Communication Networks, Association for Computer Machinery, 2017, http://porto.polito.it/2675141/2/ssl_qoe_tma_open.pdf.

Figure 16. Initial RTT Can Be Measured Anywhere



Source: Jasper, "Determining TCP Initial Round Trip Time," Packet Foo, July 15, 2014, https://blog.packet-foo.com/2014/07/determining-tcp-initial-round-trip-time.

domains that frequently impair page load times. This tool enables ISPs to cache unencrypted page elements coming from slow support domain servers and speed up page loads perceptually even when network elements are underperforming.

The insights PAIN offers are harder for ISPs to address for encrypted content, but not impossible. They would help ISPs locate domains that are persistently slow because of "hot interconnections," but additional measurement would be needed to prove that the interconnect rather than the external server is to blame.[46]

Measuring Round-Trip Time

While PAIN gathers time stamps from router log entries, other measurement approaches use self-contained monitors that passively sniff traffic. One monitoring method that is useful if not comprehensive is passive measurement of initial TCP round-trip time (RTT).

Opening a TCP "virtual circuit" connection requires a three-way handshake:

1. The initiator sends a synchronize (SYN) packet to the responder requesting a connection.

2. The responder replies with a packet acknowledging the SYN and requesting a connection from its side with another SYN.

3. The initiator acknowledges the responder's SYN.

A monitor located in some arbitrary middle can determine the RTT by counting from the initial SYN to the final acknowledgement.

RTT is an important internal metric for TCP because it controls the amount of unacknowledged data a transmitter can place on the network. Like ping, it is a measurement of the latency between two internet nodes. But unlike ping, it is an actual end-to-end measurement from the browser to the web server

Table 1. Sandvine PTS Quality of Experience Solution

Requirement	Explanation of Sandvine Solution
Ability to measure page load time, from initial request to load completion	The PTS sits in-line and inspects web-browsing traffic. The PTS is able to detect the time at which a page load is initiated and the time at which it is completed, and from those points it can determine the page load time.
Machine learning to build and regularly refresh webpage anatomy profiles	On a daily basis, using an integrated browser and a webpage parser, the Sandvine solution builds complete, detailed anatomy profiles of each page to be monitored.
Mechanism to automatically build and refresh a list of the Top N webpages	The PTS monitors actual subscriber browsing to identify the most popular webpages, refreshing the list on a daily basis.
Mechanism to allow a communication service provider (CSP) to specify particular webpages	The solution allows CSPs to define a list of pages to be monitored in addition to the empirically determined Top N list.
Ability to determine when an HTTP GET corresponds to a new page load	Through a combination of heuristics and observations, the PTS can determine which GETs correspond to new page loads and which can be ignored.
Ability to provide a comprehensive set of attributes associated with each monitored page load	For each web-browsing QoE observation provided, the PTS also provides an accompanying set of more than 20 associated attributes.
A meaningful web-browsing QoE metric	Sandvine's web-browsing QoE metric is calculated using the ITU G.1030 specification and is calibrated against real-world subscriber experiences.

Source: Sandvine, "Measuring Web Browing Quality of Experience: Requirements for Gaining Meaningful Insight," 2017, https://www.sandvine.com/resources/whitepapers/measuring-web-browsing-quality-of-experience.html.

that works its way up the protocol stack to the TCP layer. Ping does not reveal overloaded servers, but RTT does because it runs at the IP layer.

RTT is most meaningful in combination with history. Locating the capture device in the ISP footprint enables RTT for given IP addresses to be compared over time. When the latency between the browser and the capture device is known, variations in RTT to given IP addresses indicate load on the network or the end point device.

Initial RTT measurement would be useful for calibrating the measurements made by PAIN. But it should be noted that RTT is less meaningful when the web server employs a load balancer or a proxy.

Sandvine Browsing QoE System

The Sandvine Policy Traffic Switch (PTS) is an appliance used by ISPs to measure network performance.[47] PTS has a web QoE module that performs the same kinds of measurements as PAIN but on a real-time basis. The requirements statement for the product confirms the similarity of the two approaches (Table 1).

The emphasis of this tool is on features under the control of ISPs on the assumption that ISPs are the best motivated players in the ecosystem to measure performance and remediate. While ISPs have a good vantage point for observation and measurement, the assumption that they are best suited to remediate is doubtful.

SamKnows Whitebox

The SamKnows Whitebox is used in connection with test servers by 40 national regulators around the world to hold ISPs accountable for delivering on advertised speed claims.[48] It is situated in a useful location between end devices and the connection of residential and small office and home office networks to their ISPs. Effectively, the Whitebox replaces the residential internet gateway.

SamKnows has added code to the Whitebox to estimate end-to-end application performance by simulating browsers and other applications. The advantage of this approach over true monitoring is that it isolates performance factors in the user device from the measurement. While this is useful from the accountability perspective, it leaves users and developers in the dark regarding QoE. It is possible to monitor real-world conditions with a device such as the Whitebox with the proper programming.

Because the Whitebox is built on open source code, motivated researchers could write their own monitoring software for targeted research. Some of the traffic management detection tools feature customized software in a gateway device. These systems are measuring different features—generally with limited success—but they could be repurposed to measure QoE.

The Whitebox test suite currently includes tests for web browsing, Voice over Internet Protocol, video streaming, Netflix, YouTube, BBC iPlayer, and other features and applications in addition to speed tests.[49]

Tailoring Measurement to the Web's Design

The web is a system in which the party with the greatest insight—the ISP—has the least ability to exercise control. And what little control ISPs do have is intensely scrutinized by regulators, politicians, and activists.

An ideal system of measurement and remediation would involve ISPs taking web QoE measurements and sharing them with web developers and operators. A sharing arrangement would serve the interests of both parties, provided privacy rights are respected and data are not shared by parties without a legitimate interest. Under proper controls, researchers and regulators should be entitled to access anonymized data as well.

Web developers themselves have increasing stores of performance data thanks to the standards development work at W3C, so the sharing relationship could extend in both directions. Performance data do not require knowledge of payloads, as the PAIN system illustrates.

Performance data that pool the insights available to both ISPs and web developers have the potential to make the web run faster and accelerate the refinement of TCP, quick user internet connections, and similar protocols.[50] Many of the web's performance issues stem from the continued use of old technology for new tasks.

Replacing legacy parts of the internet—such as TCP and older versions of the web protocol, HTTP—with new parts better tailored to today's tasks will lower costs for websites and increase user satisfaction. However, developing, testing, and proving new protocols depends on good real-world data.

This sort of data can also help regulators wean themselves from systems of measurement that provide so little value to web users that they have effectively become political instruments. The FCC's obvious manipulation of the plateau point in broadband speeds is illustrative.

While the data show that broadband speeds above 12–15 Mbps do not allow webpages to load faster, the agency revised the figure to 25 Mbps to support a policy conclusion. This is a matter of reporting data based on a preferred policy outcome rather than developing policy on the basis of data, a cardinal sin in regulation.

The Wheeler FCC was incentivized to report misleading data on broadband networks to justify a change in the definition of broadband from 10 to 25 Mbps. This change allowed the agency to claim a low level of competition for "genuine broadband" services. This finding then supported the agency's reclassification of internet service from the lightly regulated Title I to Title II, the regulatory category tailored to telephone monopolies.

Emphasizing broadband speed has coincided with the development of faster networks. It is doubtful that measurement has promoted higher capacity networks all by itself because ISPs are motivated to upgrade speeds for business reasons. Moore's law has also enabled speeds to increase in routine equipment replacement.

Despite the minimal level of competition in the wireline ISP sector in the US, wireline ISPs face the looming threat of 5G networks taking substantial business as wireless speeds increase and customers place greater value on mobility and ease of installation than on excess performance. Making webpage QoE more visible has the potential to intensify competition between wired and wireless networks to the benefit of the public.

It can also provide guidance to web developers to improve the design of their pages and their server networks to improve performance. The internet is built on the model of multi-stakeholder collaboration. Applying this model to real-world performance measurement can make the internet experience better for everyone.

About the Author

Richard Bennett is the founder and manager of High Tech Forum, a policy initiative that educates policymakers on the technical underpinnings of tech policy issues. He is a coinventor of Wi-Fi and modern Ethernet architecture and a former visiting fellow at AEI.

© 2017 by the American Enterprise Institute. All rights reserved.

The American Enterprise Institute (AEI) is a nonpartisan, nonprofit, 501(c)(3) educational organization and does not take institutional positions on any issues. The views expressed here are those of the author(s).

Notes

1. This paper is a slightly revised version of a paper presented at TPRC 45, the Research Conference on Communications, Information and Internet Policy in Arlington, Virginia, on September 8, 2017.

2. Lucy Hazell, Peter Thompson, and Neil Davies, "A Study of Traffic Management Detection Methods & Tools," 2015, https://www.researchgate.net/publication/280714508_A_Study_of_Traffic_Management_Detection_Methods_Tools.

3. Vaibhav Bajpai and Jurgen Schonwalder, "A Survey on Internet Performance Measurement Platforms and Related Standardization Efforts," *IEEE Communications Surveys & Tutorials* 17, no. 3 (2015): 1313–41, http://ieeexplore.ieee.org/document/7076582.

4. Akamai, "State of the Internet," 2017, http://www.akamai.com/stateoftheinternet/?WT.ac=soti_banner.

5. Sandvine, *Global Internet Phenomena Report: 1H 2014*, 2014, https://www.sandvine.com/downloads/general/global-internet-phenomena/2014/1h-2014-global-internet-phenomena-report.pdf.

6. Speedtest by Ookla, "United States," July 2017, http://www.speedtest.net/global-index/united-states#fixed.

7. Cisco Systems, "Cisco Visual Networking Index: Forecast and Methodology, 2009–2014," June 2, 2010, http://www.cisco.com/en/US/solutions/collateral/ns341/ns525/ns537/ns705/ns827/white_paper_c11-481360.pdf.

8. David P. Reed, Jerome H. Saltzer, and David D. Clark, "Active Networking and End-To-End Arguments," *IEEE Network* 12, no. 3 (June 1998): 69–71, http://ieeexplore.ieee.org/document/690972.

9. Jim Rapoza, "The Web Is Getting Slower," Aberdeen Essentials, November 5, 2014, http://www.aberdeenessentials.com/techpro-essentials/the-web-is-getting-slower; Kalev Leetaru, "Why the Web Is So Slow And What It Tells Us About the Future of Online Journalism," *Forbes*, February 6, 2016, https://www.forbes.com/sites/kalevleetaru/2016/02/06/why-the-web-is-so-slow-and-what-it-tells-us-about-the-future-of-online-journalism; and Hope King, "The Web Is Getting Slower," CNNMoney, June 16, 2015, http://money.cnn.com/2015/06/16/technology/web-slow-big/index.html.

10. Richard Bennett, "Open Internet Orders Degrade Internet Improvement," High Tech Forum, June 19, 2017, http://hightechforum.org/open-internet-orders-degrade-internet-improvement; and Federal Communications Commission, Office of Engineering and Technology and Office of Strategic Planning & Policy Analysis, "Measuring Fixed Broadband Report 2016," September 29, 2016, https://www.fcc.gov/reports-research/reports/measuring-broadband-america/measuring-fixed-broadband-report-2016.

11. Akamai, "State of the Internet Connectivity Visualization," August 2017, https://www.akamai.com/us/en/about/our-thinking/state-of-the-internet-report/state-of-the-internet-connectivity-visualization.jsp.

12. Federal Communications Commission, Office of Engineering and Technology and Consumer and Governmental Affairs Bureau, "Measuring Broadband America," 2016, https://www.fcc.gov/general/measuring-broadband-america.

13. Federal Communications Commission, "Measuring Fixed Broadband Report 2016."

14. Ibid., Figure 1.

15. Speedtest by Ookla, "United States Speedtest Market Report," August 3, 2016, http://www.speedtest.net/reports/united-states/; Richard Bennett, *G7 Broadband Dynamics: How Policy Affects Broadband Quality in Powerhouse Nations*, American Enterprise Institute, November 4, 2014, http://www.aei.org/publication/g7-broadband-dynamics-policy-affects-broadband-quality-powerhouse-nations/; and Speedtest by Ookla, "United States."

16. Rapoza, "The Web Is Getting Slower."

17. Leetaru, "Why the Web Is So Slow and What It Tells Us About the Future of Online Journalism."

18. King, "The Web Is Getting Slower."

19. Akamai, "Real User Monitoring for Improved Web Experience," 2017, https://www.akamai.com/us/en/resources/real-user-monitoring.jsp.

20. David Belson, "State of the Internet Metrics: What Do They Mean?," *Akamai Blog*, February 5, 2015, https://blogs.akamai.com/2015/02/state-of-the-internet-metrics-what-do-they-mean.html.

21. Akamai, "State of the Internet Connectivity Visualization."

22. Hazell, Thompson, and Davies, "A Study of Traffic Management Detection Methods & Tools."

23. Bajpai and Schonwalder, "A Survey on Internet Performance Measurement Platforms and Related Standardization Efforts."

24. Hazell, Thompson, and Davies, "A Study of Traffic Management Detection Methods & Tools," 46.

25. Federal Communications Commission, "Measuring Broadband America," 6.

26. Federal Communications Commission, Office of Engineering and Technology and Consumer and Governmental Affairs Bureau, "Measuring Broadband America Fixed Broadband Report—2015," December 30, 2015, https://www.fcc.gov/reports-research/reports/measuring-broadband-america/measuring-broadband-america-2015.

27. Federal Communications Commission, "Archive of Released Broadband Progress Notices of Inquiry," March 27, 2014, https://www.fcc.gov/encyclopedia/archive-released-broadband-progress-notices-inquiry.

28. Robert B. Miller, "Response Time in Man-Computer Conversational Transactions," ACM Press, 1968, 267, http://theixdlibrary.com/pdf/Miller1968.pdf.

29. Lauren Younger, "How to Test Smarter with RUM-Based Performance Testing," SOASTA, September 1, 2016, https://www.soasta.com/blog/rum-based-web-performance-testing/.

30. Akamai, "The State of Online Retail Performance," 2017, https://www.soasta.com/wp-content/uploads/2017/04/State-of-Online-Retail-Performance-Spring-2017.pdf.

31. Ibid.

32. HTTP Archive, "Trends," accessed August 13, 2017, http://httparchive.org/trends.php.

33. Ghostery, "Ghostery," https://www.ghostery.com.

34. Anecdotally, I have observed one site with as many as 130 trackers: op-eds on the LA Times site.

35. HTTP Archive, "Trends"; and Akamai, "State of the Internet Connectivity Visualization."

36. Akamai, "State of the Internet"; Federal Communications Commission, "Measuring Broadband America"; and John Stevens, "How Slow Is Too Slow in 2016?," Webdesigner Depot, February 12, 2016, https://www.webdesignerdepot.com/2016/02/how-slow-is-too-slow-in-2016.

37. Mark Allman, Vern Paxson, and Ethan Blanton, "TCP Congestion Control," Network Working Group, September 2009, https://tools.ietf.org/html/rfc5681.

38. Neil Cardwell et al., "TCP BBR Congestion Control Comes to GCP—Your Internet Just Got Faster," *Google Cloud Platform Blog*, July 20, 2017, https://cloudplatform.googleblog.com/2017/07/TCP-BBR-congestion-control-comes-to-GCP-your-Internet-just-got-faster.html.

39. Rob Pegoraro, "Slow Internet Downloads? Look to Wi-Fi First, Then Blame Your ISP," *USA Today*, July 23, 2017, https://www.usatoday.com/story/tech/columnist/2017/07/23/slow-internet-downloads-look-wi-fi-first-then-blame-your-isp/501173001; Whitson Gordon, "Top 10 Ways to Deal with a Slow Internet Connection," Lifehacker, June 22, 2013, http://lifehacker.com/top-10-ways-to-deal-with-a-slow-internet-connection-514138634; and Bradley Mitchell, "Tips for Troubleshooting Slow Internet Connections at Home," Lifewire, April 7, 2017, https://www.lifewire.com/troubleshooting-slow-internet-connection-818125.

40. W3C, "Web Performance Current Status," W3C Standards, https://www.w3.org/standards/techs/performance#w3c_all.

41. Ilya Grigorik, James Simonsen, and Jatinder Mann, "High Resolution Time Level 2," W3C Standards, August 3, 2017, https://www.w3.org/TR/2017/CR-hr-time-2-20170803/.

42. Arvind Jain et al., "Resource Timing Level 1," W3C Standards, March 30, 2017, https://www.w3.org/TR/resource-timing-1/.

43. HTTP Archive, "Trends."

44. Martino Trevisan, Idilio Drago, and Marco Mellia, "PAIN: A Passive Web Speed Indicator for ISPs," ACM SIGCOMM Workshop on QoE-based Analysis and Management of Data Communication Networks, Association for Computer Machinery, 2017, http://porto.polito.it/2675141/2/ssl_qoe_tma_open.pdf.

45. Ibid., 1.

46. William B. Norton, *The Internet Peering Playbook: Connecting to the Core of the Internet* (Palo Alto, CA: DrPeering Press, 2011), 98.

47. Sandvine, "Measuring Web Browing Quality of Experience: Requirements for Gaining Meaningful Insight," 2017, https://www.sandvine.com/resources/whitepapers/measuring-web-browsing-quality-of-experience.html.

48. SamKnows, "The Whitebox," https://www.samknows.com/products.

49. SamKnows, "SamKnows Test Methodology: Methodology and Technical Information Relating to the SamKnows Testing Platform," February 2015, https://files.samknows.com/~whitepapers/SQ301-005-EN-Test-Suite-Whitepaper-4.pdf.

50. Internet Engineering Task Force, "QUIC," February 27, 2017, https://datatracker.ietf.org/wg/quic/about/.